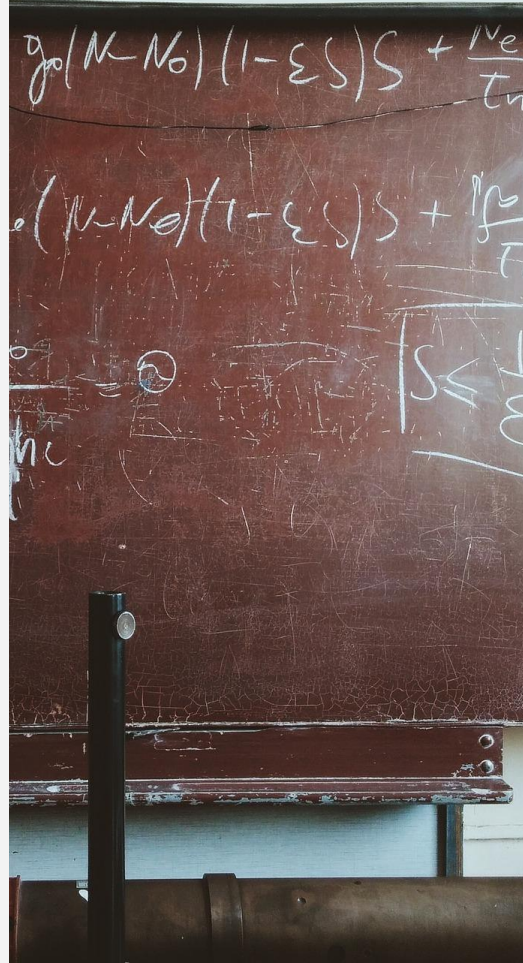


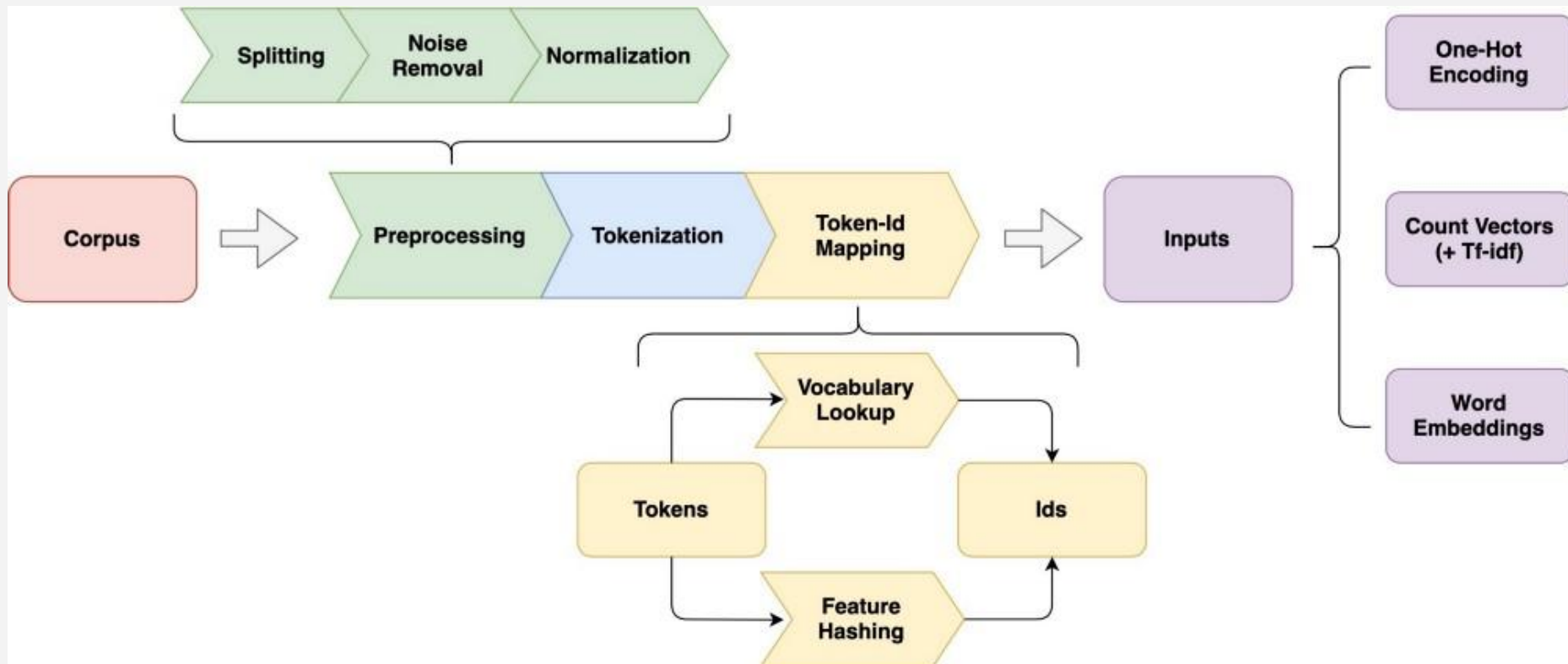
Análise de Dados

01

PRÉ-PROCESSA MENTO

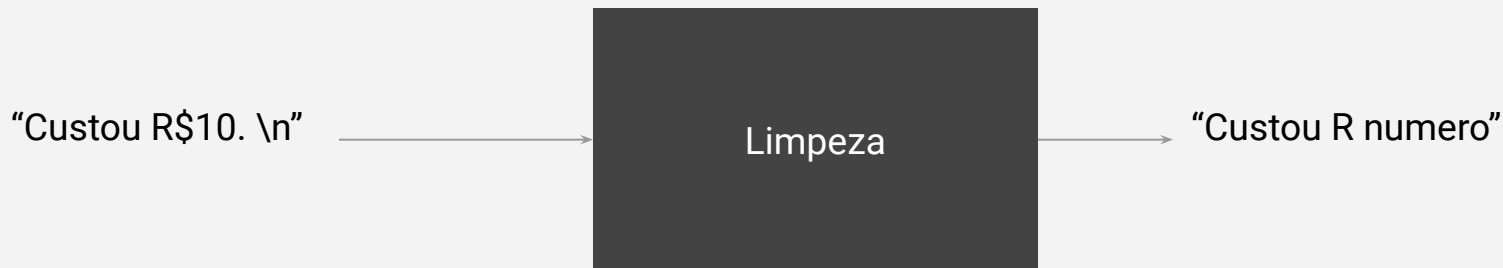


PRE PROCESSAMENTO



Limpeza

- Remoção de Caracteres especiais
- Remoção de Números
- Remoção de Hyperlinks
- Pontuação



Stop Word

Stopword são palavra que possui apenas significado sintático dentro da sentença, porém não traz informações relevantes sobre o seu sentido.

Exemplo: “de”, “da”, “a”, “o”

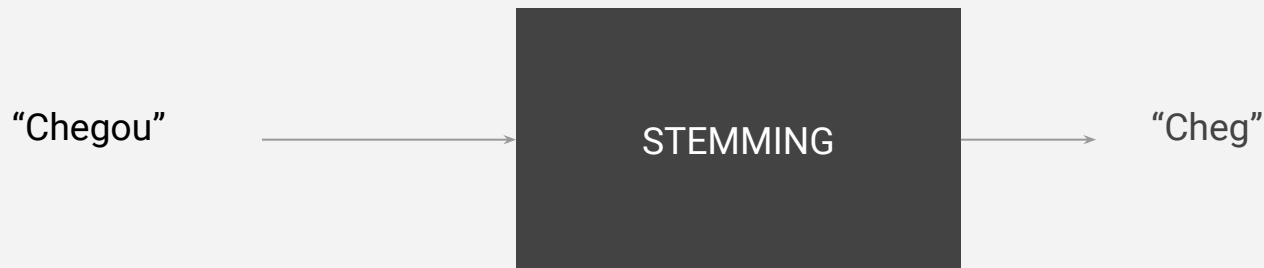


Stemming

Reduz os termos ao seu radical, removendo afixos e vogais temáticas.

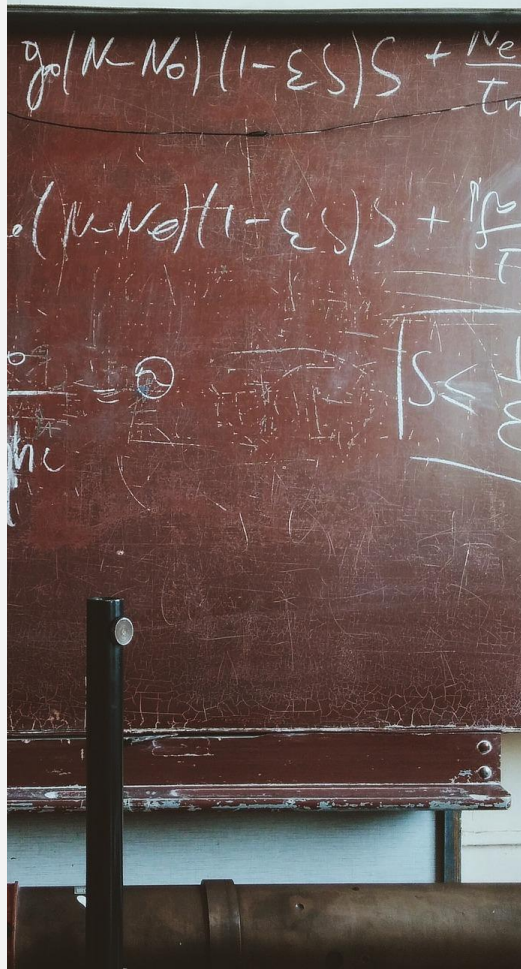
Exemplo:

- Entregar se torna Entreg
- Copiar se torna Copi



02

REPRESENTAÇÃO DE TEXTO



Bag of Word

é o melhor dos melhores.

→ ENTRADA

é o melhor dos melhores bom legal

→ DICIONÁRIO

1 1 1 1 1 0 0

→ REPRESENTAÇÃO

Contagem

é o melhor dos melhores.

—————→ ENTRADA

é o melhor dos melhores bom legal

—————→ DICIONÁRIO

1 2 1 2 1 0 0

—————→ REPRESENTAÇÃO

TD-IDF

Um problema com a frequência de palavras de pontuação é que palavras altamente frequentes começam a dominar no documento (por exemplo, pontuação maior), mas podem não conter tanto "conteúdo informativo" para o modelo como palavras mais raras, mas talvez de domínio específico.

TD

Term Frequency



Bag of Word

IDF

Inverse Document Frequency



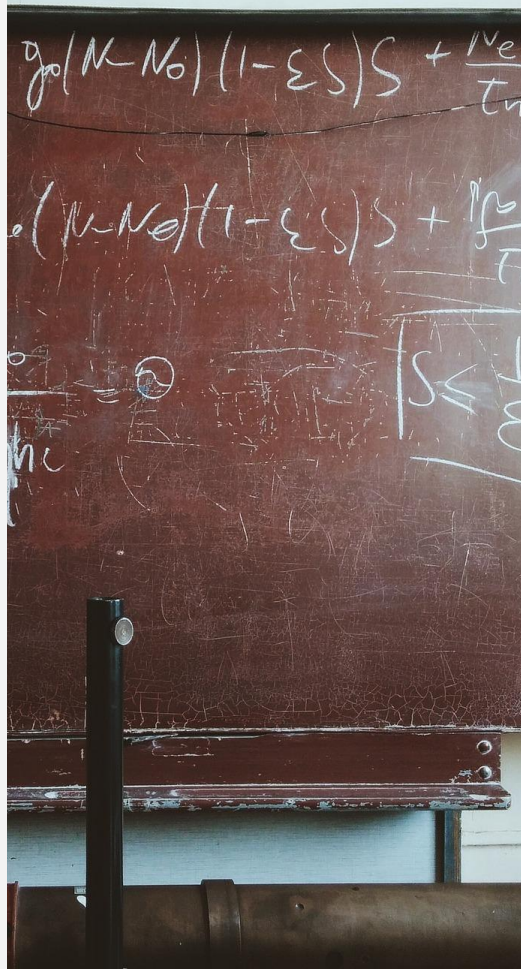
Pontuação da palavra
aparece em outros
documentos

TD-IDF

Term	Review 1	Review 2	Review 3	IDF
This	1	1	1	0.00
movie	1	1	1	0.00
is	1	2	1	0.00
very	1	0	0	0.48
scary	1	1	0	0.18
and	1	1	1	0.00
long	1	0	0	0.48
not	0	1	0	0.48
slow	0	1	0	0.48
spooky	0	0	1	0.48
good	0	0	1	0.48

03

ANÁLISE DE SENTIMENTO



Polaridade

- “Gostei do produto chegou rápido.” (Positivo)
- “Gostei do produto, mas demorou para chegar.” (Neutro)
- “Produto horrível e demorou para entregar.” (Negativo)

Força do Sentimento

- “Gostei do produto chegou rápido.” (0.99)
- “Gostei do produto, mas demorou para chegar.” (0.2)
- “Produto horrível e demorou para entregar.” (-0.99)

Emoções

- “Gostei do produto chegou rápido.” (felicidade)
- “Gostei do produto, mas demorou para chegar.” (felicidade, frustração)
- “Produto horrível e demorou para entregar.” (frustração, raiva)

Baseado em Regras

Regras

Dicionário

Baseadas em Recursos

Machine
Learning

Deep Learning